Extensibility: Customize HDInsight Clusters Using Script Actions

Extensibility, customize HDInsight clusters using script actions.
In this video we'll go over HDInsight cluster customization options and how to use a script action during cluster creation.
Persisted scripts are scripts that are applied to worker nodes and will be run automatically on new nodes created when scaling up a cluster.
A persisted script might also apply changes to another node type such as a headnode, but form a functionality perspective the only reason to persist a script is so it applies to a new worker nodes created when a clusters scaled out.
Ad hoc scripts are not persisted, however you can subsequently promote an ad hoc script to persisted script or demote a persisted script to an ad hoc script.
A script action is simply a Bash script that you provide a URL to, and parameters for.
And it is then run on the HDInsight cluster nodes.
Script actions must complete within 60 minutes, or they will time out.
During cluster provisioning, the script is run concurrently with other setup and configuration processes.
Competition for resources such as CPU time or network bandwidth may cause the script to take longer to finish than it does in your development environment.
During cluster creation, you can specify multiple script actions that are invoked in the order in which they were specified.
Script actions used during cluster creation are slightly different from script actions run on an existing cluster.
The script is automatically persisted.
A failure in the script can cause the cluster creation process itself to fail.
The script is run while HD Insight is being configured.
At this stage, the script is run in parallel on all the specified nodes in the cluster and with root privileges on the nodes.
During cluster creation, you can specify multiple script actions that are invoked in the order in which they were specified.
These are the best practices to be followed with regard to script actions.
Check for the Hadoop version.
Only HDInsight version 3.1 with Hadoop 2.4 and above support using script actions to install custom components on a cluster.
Provides stable links to script resources.
Users should make sure that all of the scripts and other artifacts used in the customization of a cluster remain available throughout the lifetime of the cluster and that the versions of these files did not change for the duration.
Ensure that the cluster customization script is item potent, you must expect that the nodes of an HDInsight

cluster will be reimaged during the cluster lifetime.
The cluster customization script is run whenever cluster
is reimaged.
Configure the custom components to use Azure Blob Storage.
The custom components that you install on the cluster
nodes might have a default configuration
to use Hadoop distributive file system, HTFS storage.
These are the components that can be installed on
HD insight cluster's using scripts providing by HD insight.
Script Action Scripts can be used from the Azure portal,
Azure PowerShell, Azure CLI, or the HDN site.NET SDK.
Script Action Scripts can be used from the listed
tools depending upon whether this script is applied
while creating or running the cluster.
This section provides examples on the different ways you can
use script actions when creating an HDInsight cluster from
the Azure portal using an ARM template,
using PowerShell CMDlets, and using the .NET SDK.
This is how to use Script Actions when creating an HD
inside cluster from the Azure portal.
Start creating a cluster under Optional Configuration, for
the Script Actions blade.
Click Add Script Action to provide details about
the script action.
Add property details,
press enter to add more than one script action to install
multiple components on the cluster.
Click Select to save the configuration and
configure with Cluster creation.
To create a Cluster using Azure Resource Manager create an ARM
template, and copy the template to a location on your computer.
Then start Azure PowerShell, and log in to your Azure account.
If you have multiple subscriptions,
provide the subscription ID you wish to be use for deployments.
If you do not have an existing resource group,
create a new resource group.
Then provide the name of the resource group and
location that you need for your solution.
To create a new deployment for
your resource group, run the New-AzureRmResourceGroupDeploy-
ment command, and provide the necessary parameters.
The parameters will include a name for your deployment,
the name of your resource group,
and the path or URL to the template you created.
To use the Add-AzureRmHDInsightScriptAction
cmdlet to invoke scripts by using Script Action to customize
a cluster, open the Azure PowerShell console, and log in
to your Azure subscription and declare PowerShell variables.
Specify the configuration values, such as nodes
in the cluster, and the default storage to be used. Use

Add-AzureRmHDInsightScriptAction
cmdlet to invoke the script.
Set the admin https user for the cluster.
Then set the SSH credentials and create the cluster.
The HD Insight.net SDK provides .NET client
libraries that make it easier to work with HD Insight from
a .NET Framework application to create Linux based clusters.
You can apply script actions to a running HD Insight cluster
from the Azure portal using the PowerShell commandlets
using the cross platform Azure CLI.
And using the .NET SDK.
Follow the steps listed in this slide to apply a script action
to a running cluster from the Azure portal.
Open the Azure PowerShell console, and log into your Azure
subscription and declare PowerShell variables.
Use the command given in this slide to apply the script
to the cluster.
Open a shell session, terminal, command prompt,
or other command line for your system and use the command given
in this slide to switch to Azure Resource Manager mode.